

Applying Rough Sets Reduction Techniques to the Construction of a Fuzzy Rule Base for Case Based Reasoning

Florentino Fdez-Riverola¹, Fernando Díaz¹, and Juan M. Corchado²

¹ Dept. Informática, University of Vigo, Escuela Superior de Ingeniería Informática, Edificio Politécnico, Campus Universitario As Lagoas s/n, 32004, Ourense, Spain
{riverola, fdiaz}@uvigo.es

² Dept. de Informática y Automática, University of Salamanca,
Plaza de la Merced s/n, 37008, Salamanca, Spain
corchado@usal.es

Abstract. Early work on Case Based Reasoning reported in the literature shows the importance of soft computing techniques applied to different stages of the classical 4-step CBR life cycle. This paper proposes a reduction technique based on Rough Sets theory that is able to minimize the case base by analyzing the contribution of each feature. Inspired by the application of the minimum description length principle, the method uses the granularity of the original data to compute the relevance of each attribute. The rough feature weighting and selection method is applied as a pre-processing step previous to the generation of a fuzzy rule base that can be employed in the revision phase of a CBR system. Experiments using real oceanographic data show that the proposed reduction method maintains the accuracy of the employed fuzzy rules, while reducing the computational effort needed in its generation and increasing the explanatory strength of the fuzzy rules.

1 Introduction and Motivation

Case-Based Reasoning (CBR) systems solve problems by reusing the solutions to similar problems stored as cases in a case base [1] (also known as memory). However, these systems are sensitive to the cases present in the case memory and often its accuracy depend on the significance of these stored cases. Therefore, in CBR systems it is always important to reduce noisy cases in order to achieve a good generalisation accuracy.

Case Based Reasoning is an application area where soft computing tools have had a significant impact during the past decade [2]. These soft computing techniques (fuzzy logic, artificial neural networks, genetic algorithms and rough sets, mainly) work in parallel for enhancing the problem-solving ability of each other [3].

In this paper we propose a reduction technique based on Rough Set theory applied to the revision stage of CBR systems. The proposed method was introduced into our Case-Based forecasting platform called FSfRT, specialized in making predictions for

changing environments. Presently, the FSfRT platform is able to combine artificial neural networks and fuzzy logic in the framework of a CBR system.

The paper is structured as follows: section 2 introduces the Rough Set theory grounding; section 3 details the proposed Rough Set reduction technique; section 4 describes the Case-Based Reasoning platform used in this study; section 5 exposes the test bed of the experiments and the results obtained; finally, section 6 presents the conclusions and further work.

2 Rough Set Theory

Rough Set theory, proposed by Pawlak [4,5], is an attempt to provide a formal framework for the automated transformation of data into knowledge. It is based on the idea that any inexact concept (for example, a class label) can be approximated from below and from above using an indiscernibility relationship. Pawlak [6] points out that one of the most important and fundamental notions to the Rough Set philosophy is the need to discover redundancy and dependencies between features.

The main advantages of Rough Set theory are that it: (i) provides efficient algorithms for discovering hidden patterns in data; (ii) identifies relationships that would not be found while using statistical methods; (iii) allows the use of both qualitative and quantitative data; (iv) finds the minimal sets of data that can be used for classification tasks; (v) evaluates the significance of data and (vi) generates sets of decision rules from data.

2.1 Basic Concepts and Definition

Briefly, the relevant Rough Set terminology is stated below. An information system is a pair $S = \langle U, A \rangle$, where U is a non-empty and finite set, called the universe, and A is a non-empty, finite set of attributes (or features). An equivalence relation, referred to as indiscernibility relation, is associated with every subset of attributes $P \subseteq A$. This relation is defined as:

$$IND(P) = \{(x, y) \in U \times U : \text{for every } a \in P, a(x) = a(y)\}. \quad (1)$$

Given any subset of features P , any concept $X \subseteq U$ can be defined approximately by the employment of two sets, called lower and upper approximations. The lower approximation, denoted by $\underline{P}X$, is the set of objects in U which can be certainty classified as elements in the concept X using the set of attributes P , and is defined as follows:

$$\underline{P}X = \cup\{Y \in U / IND(P) : Y \subseteq X\}. \quad (2)$$

The upper approximation, denoted by $\overline{P}X$, is the set of elements in U that can be possibly classified as elements in X , formally:

$$\overline{P}X = \cup\{Y \in U / IND(P) : Y \cap X \neq \emptyset\}. \quad (3)$$

The degree of dependency of a set of features P on a set of features R is denoted by $\gamma_R(P)$, $0 \leq \gamma_R(P) \leq 1$, and is defined as:

$$\gamma_R(P) = \frac{\text{card}(\text{POS}_R(P))}{\text{card}(U)}. \quad (4)$$

where

$$\text{POS}_R(P) = \bigcup_{X \in U / \text{IND}(P)} \underline{RX}. \quad (5)$$

$\text{POS}_R(P)$ contains the objects of U which can be classified as belonging to one of the equivalence classes of $\text{IND}(P)$, using only features from the set R . If $\gamma_R(P) = 1$, then R functionally determines P .

Various extensions have been defined from the basic model proposed by Pawlak. Among these extensions stands out the Variable Precision Rough Set model (VPRS) [7] which is a generalisation that introduces a controlled degree of uncertainty within its formalism. This degree is established by an additional parameter ϕ .

2.2 Rough Sets as Reduction Technique

A major feature of the Rough Set theory is to find the minimal sets of data that can be used for classification tasks. In this sense, the notions of core and reduct of knowledge are fundamental for reducing knowledge preserving information. After stating the formal definitions of these concepts, it is outlined the reduction process proposed by the methodology.

P is an independent set of features if there does not exist a strict subset P' of P such that $\text{IND}(P) = \text{IND}(P')$. A set $R \subseteq P$ is a reduct of P if it is independent and $\text{IND}(R) = \text{IND}(P)$. Each reduct has the property that a feature can not be removed from it without changing the indiscernibility relation. Many reducts for a given set of features P may exist. The set of attributes belonging to the intersection of all reducts of P is called the core of P :

$$\text{core}(P) = \bigcap_{R \in \text{Reduct}(P)} R. \quad (6)$$

An attribute $a \in P$ is indispensable if $\text{IND}(P) \neq \text{IND}(P \setminus \{a\})$. The core of P is the union of all the indispensable features in P .

The reduction technique stated by the methodology is specially suitable for reducing decision tables. A decision table is an information system of the form $S = \langle U, A \cup \{d\} \rangle$, where $d \notin A$ is a distinguished attribute called the decision attribute or class attribute. The elements of the set A are referred to as condition attributes. A decision table is a classifier that has as its internal structure a table of labelled instances. Given a novel instance, the classification process is based on the search of all matching in-stances in the table. If no matching instances are found, unknown is

returned; otherwise, the majority class of the matching instances is returned (there may be multiple matching instances with conflicting labels). The indispensable attributes, reducts, and core can be similarly defined relative to a decision attribute or output feature. The precise definitions of these concepts can be found in Pawlak's book on Rough Sets [5].

At this point, it is very important to use the classification rules (given by a decision table) with the minimal effort, and therefore, the simplification of decision tables is of primary importance. The simplification process comprises two fundamental tasks. On the one hand, reduction of attributes consists of removing redundant or irrelevant attributes, without losing any essential classification information. The computation of the reducts for the condition attributes relative to the decision attribute is carried out to achieve this goal. On the other hand, reduction of attribute values is related to the elimination of the greatest number of condition attribute values, maintaining also the classificatory power.

3 Feature Subset Selection Using Rough Sets

The computation of the reducts and the core of the condition attributes from a decision table is a way of selecting relevant features. It is a global method in the sense that the resultant reduct represents the minimal set of features which are necessary to maintain the same classificatory power given by the original and complete set of attributes. A straighter manner for selecting relevant features is to assign a measure of relevance to each attribute and choose the attributes with higher values.

In the Rough Set framework, the natural way to measure the prediction success is the degree of dependency defined above. However, [8] have shown the weakness of this measure in order to assess an estimation of the predictive accuracy of a set of condition attributes Q with regard to a class attribute d . To overcome this deficiencies, [9] define the notion of rough entropy. Based on this notion and its adaptation to the VPRS model (in order to exploit more efficiently the knowledge that is provided for the observations in the boundary region or the uncertain area of the universe), we have defined a coefficient that allows to assess the significance of an attribute within a set of attributes [10]. The significance of an attribute $a \in Q$ is defined in a way that its value is greater when the removal of this attribute leads to a greater diminution of the complexity of the hypothesis $Q \setminus \{a\}$, and simultaneously, to a lesser loss of accuracy of the hypothesis. Implicitly, the underlying principle used to evaluate the relevance of an attribute in this way is the Minimum Description Length Principle (MDLP) [11].

The associated complexity of a given set of condition attributes Q can be evaluated through the entropy of the partition $U / IND(Q)$, which will be denoted by $H(Q)$. On the other hand, the conditional rough entropy $H_\phi(d | Q)$ can be used to evaluate the accuracy that is achieved when the condition attributes Q are used to predict the value of the condition attribute d . Therefore, the formal definition of the ϕ -rough entropy, denoted by $RH_\phi(d | Q)$, is given by the following expression:

$$\begin{aligned}
RH_\phi(Q, d) &= H(Q) + H_\phi(d \mid Q) = H(Q) + \\
&\left\{ \{1 - \gamma_{Q,\phi}(d)\} \log_2 |U| + \sum_{x_i \in \text{POS}_{Q,\phi}(d)} \frac{|X_i|}{|U|} \log_2 \frac{|X_i|}{|U|} \right\}. \\
&= \{1 - \gamma_{Q,\phi}(d)\} \log_2 |U| - \sum_{x_i \in \text{POS}_{Q,\phi}(d)} \frac{|X_i|}{|U|} \log_2 \frac{|X_i|}{|U|}
\end{aligned} \tag{7}$$

where X_i represents each one of the classes of the partition $U / \text{IND}(Q)$, the set $\text{POS}_{Q,\phi}(d)$ is the positive region of Q with regard to the decision attribute d , and $\gamma_{Q,\phi}(d)$ is the degree of dependence of attribute d on the set of attributes Q .

Then, the ϕ -significance of a condition attribute, $a \in Q$, with regard to the decision attribute d , denoted by $\sigma_{a,\phi}(Q, d)$, is defined as the variation that the ϕ -rough entropy suffers when the considered attribute is dismissed from Q . Namely, it is computed the term $\Delta_a RH_\phi(Q, d)$, given by the difference between $RH_\phi(Q, d)$ and $RH_\phi(Q \setminus \{a\}, d)$. Formally,

$$\begin{aligned}
\sigma_{a,\phi}(Q, d) &= \Delta_a RH_\phi(Q, d) = RH_\phi(Q, d) - RH_\phi(Q \setminus \{a\}, d) \\
&= \{H(Q) - H(Q \setminus \{a\})\} - \{H_\phi(d \mid Q \setminus \{a\}) - H_\phi(d \mid Q)\}
\end{aligned} \tag{8}$$

Fig. 1 provides a concise description of the algorithm that selects a subset of relevant features using the significant ϕ -rough coefficient to evaluate the relevance of a feature. The proposed algorithm for selecting relevant features is described according to the view proposed by [12]. These authors state that a convenient paradigm for viewing feature selection methods is that of heuristic search, with each state in the search space specifying a subset of the possible features. Following Blum and Langley viewpoint the four basic issues that characterise this method are:

- The starting point in the space, which in turn influences the direction of search and the operators used to generate successor states. The proposed algorithm starts with all attributes and successively removes them (lines 1 and 15, respectively). This approach is known as backward elimination.
- The organisation of the search. Any realistic approach relies on a greedy method to traverse the space considering that an exhaustive search is impractical. At each point in the search, the proposed algorithm considers all local changes, namely, it evaluates the significance of each attribute of the current set of attributes (loop for).
- The strategy used to evaluate alternative subsets of attributes. In this paper, the variation of the normalised ϕ -rough entropy has been chosen for this purpose. Specifically, at each decision point the next selected state is that one which results from removing the attribute with the least significant ϕ -rough coefficient (line 10).
- A criterion for halting the search. In the algorithm, the criterion for halting is that the difference between the degree of dependency at initial state and the current one (both with respect to the decision) do not exceed a predefined threshold (line 14).

```

Function FEATURE_SUBSET_SELECTION (input: decisionTable,  $\phi$ , output: outputFeatures)
{
00  begin
01  outputFeatures  $\leftarrow$  Features(decisionTable) /* the starting point is the complete set of input features */
02   $\gamma_{initial} \leftarrow$  CoefDependence(decisionTable, outputFeatures, TargetFeature(decisionTable),  $\phi$ )
03  step  $\leftarrow$  0; haltCriterion  $\leftarrow$  FALSE
04  while  $\neg$ haltCriterion do /* while it is not satisfied criterion of halt, remove features */
05      haltCriterion  $\leftarrow$  TRUE; step  $\leftarrow$  step + 1
06       $\sigma_{max} \leftarrow$  1.0;
07      for each feature  $f \in$  outputFeatures do /* select the most irrelevant feature */
08           $\sigma_f \leftarrow$  CoefSignificance(decisionTable, f, outputFeatures, TargetFeature(decisionTable),  $\phi$ )
09           $\gamma \leftarrow$  CoefDependence(decisionTable, outputFeatures \ {f}, TargetFeature(decisionTable),  $\phi$ )
10          if (  $\sigma_f < \sigma_{max}$  ) then
11               $\gamma_{current} \leftarrow \gamma$ 
12              feature_to_remove  $\leftarrow$  f
13           $\gamma_{threshold} \leftarrow \gamma_{initial} * (1.0 - \text{DecrementFactor}(\text{step}))$ ;
14          if (  $\gamma_{current} \geq \gamma_{threshold}$  ) then
15              outputFeatures  $\leftarrow$  outputFeatures \ {feature_to_remove} /* remove feature */
16              haltCriterion  $\leftarrow$  FALSE
17  end
}

```

Fig. 1. Algorithm for feature subset selection

4 Description of the FSfRT Platform

The study described in this paper was carried out in the context of the FSfRT platform. FSfRT is a structured hybrid system that can employ several soft computing techniques in order to accomplish the 4-steps of the classical CBR life cycle [1]. This section covers two main points: (i) details the architecture of the FSfRT platform and (ii) introduces the use of Rough Sets inside the whole system.

4.1 FSfRT Platform Architecture

The FSfRT platform is an extension of a previous successful system [13] able to make predictions of red tides (discolourations caused by dense concentrations of microscopic sea plants, known as phytoplankton). The FSfRT platform allows us to combine several soft computing techniques in order to test their suitability working together to solve complex problems. The core and the interfaces of FSfRT have been coded in Java language and new capabilities are being developed. The general idea is to have different programmed techniques able to work separately and independently in co-operation with the rest. The main goal is to obtain a general structure that could change dynamically depending on the type of problem. Fig. 2 shows a schematic view of the system.

On the left of Fig. 2, it is shown the core of the platform that is composed by a KAM (*Knowledge Acquisition Module*). The KAM is able to store all the information needed by the different techniques employed in the construction of a final CBR system. In the retrieve and reuse stages, several soft computing techniques can be used [2,3], while in the revise stage, our platform employs a set of TSK fuzzy systems [14] in order to perform the validation of the initial solution proposed by the system.

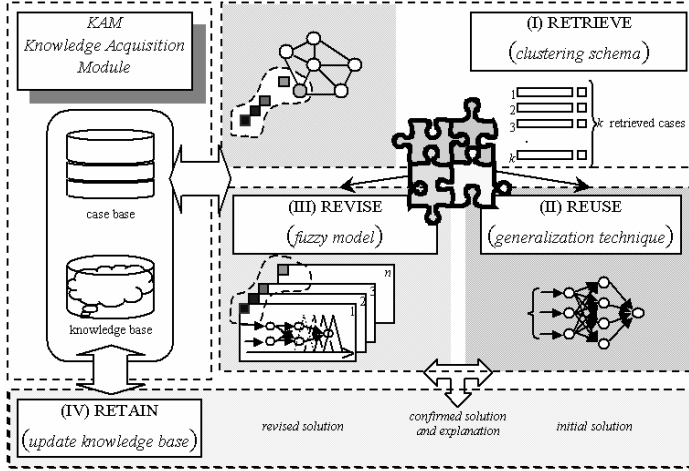


Fig. 2. FSfRT platform architecture

Our aim in this work, is to perform a feature subset selection step before the induction process carried out by the fuzzy revision method. The aim of this proposal is to reduce the original set of attributes and therefore decrease the computational effort for the generation of the different fuzzy models.

4.2 Rough Sets Inside the FSfRT Platform

Fig. 3 shows the meta-level process when incorporating the Rough Sets as a pre-processing step before the generation of the fuzzy revision subsystem.

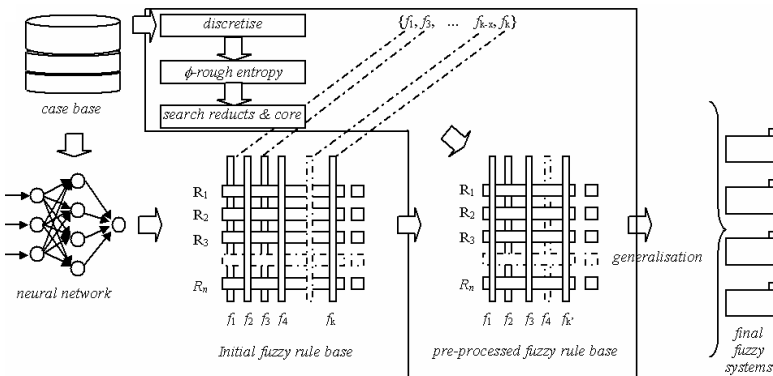


Fig. 3. Rough Set pre-processing step

For details related to the construction of the fuzzy systems starting from a Radial Basis Function neural network see [14]. The Rough Set process described here

involves the first step in the generation of the initial fuzzy system and it is divided into three phases:

The first one discretises the cases stored in the case base. It is necessary in order to find the most relevant information using the Rough Set theory. The second one uses the significant ϕ -rough coefficient to select a subset of relevant features as described in section 3 (see Fig. 1). Finally, the last phase searches for reducts and core of knowledge from the features selected in the previous phase, as explained in section 2.

The motivation of including the second phase is that the computation of reducts is a blind technique, where several combinations of a sufficient number of irrelevant features can become a reduct. The pre-selection of features leads to reducts with a lesser complexity and a higher predictive accuracy.

5 Empirical Study

In order to evaluate the proposed method, we use a biological database composed by several physical variables (temperature, PH, oxygen, salinity, etc.) measured at distinct depths and belonging to different monitoring points of the north west coast of the Iberian Peninsula. These data values are complemented with data derived from satellite images stored separately. The satellite image data values are used to generate cloud and superficial temperature indexes. The whole memory of the system consists on approximately 6300 cases, each one represented as a feature vector that holds 56 attributes.

The FSfRT platform was configured to use the same techniques as in our previous work [14], where the fuzzy revision method was successfully tested: (i) a Growing Cell Structure (GCS) neural network as retrieval method, (ii) a Radial Basis Function (RBF) neural network for the reuse step and (iii) the aforementioned set of TSK fuzzy systems working as the revision mechanism. Specific information about these techniques and its integration inside the CBR life cycle can be found in [15]. The main goal of the previous work was to develop a forecasting biological system capable of predicting the concentration of diatoms (a type of single-celled algae) in different water masses.

Although the experiments carried out in [14] showed the effectiveness and the straightforward improvement of the proposed fuzzy revision method over other approaches, some issues remained unsolved in order to deploy the application for real use. Concisely, the main drawbacks of the tested method were: (i) the time needed for generating each one of the TSK fuzzy models and (ii) the explanatory complexity of the fuzzy rules used for the final solution proposed by the system.

In order to solve these problems maintaining at the same time the accuracy level, we have proposed in this paper a feature subset selection algorithm based on Rough Set theory. As we can see in Table 1, several ϕ values have been tested in order to obtain the most accurate set of representative features defining each problem case. For the current domain of diatoms forecasting, the optimal number of features was 12 ($\phi = 0.01$), corresponding to the physical magnitudes measured with a smaller level of depth and those generated from satellite images.

A crucial aspect in this experiment is the accuracy level of the Rough Set based revision subsystem and its comparison with the original one. Starting from the error

Table 1. Selected features depending on the value of the parameter ϕ

parameter ϕ	number of selected features
0.0	16
0.01	12
0.025	10
0.05	8
0.1	8

series generated by the different models, the Kruskal-Wallis test has been carried out. Since the P-value is less than 0.01, there is a statistically significant difference among the models at the 99.0% confidence level. Fig. 4 shows a multiple comparison procedure (Mann-Whitney) used to determine which models are significantly different from the others. The experiments were made with a data set of 448 cases randomly taken from the case base. It can be seen that the CBR with TSK fuzzy revision subsystem (CBR TSK) presents statistically significant differences with the rest of the models, whilst it is as accurate as the simplified method presented here (CBR ϕ (TSK)).

	CBR ϕ (TSK)	CBR TSK	RBF	RBF + GCS	ARIMA	Quadratic Trend	Moving Average	Simp. Exp. Smooth	Lin. Exp. Smooth
CBR ϕ (TSK)	=								
CBR TSK	*	*							
RBF	*	*	=						
RBF+GCS	*	*	*	*					
ARIMA	*	*	*	*	*				
Quadratic Trend	*	*	*	*	*	*			
Moving Average	*	*	*	*	*	*	*		
Simp. Exp. Smooth.	*	*	=	=	*	*	=	*	
Lin. Exp. Smooth.	*	*	=	=	*	*	=	=	*

Fig. 4. Mann-Whitney test carried out between each pair of models

The time consumed in the execution of the pre-processing step plus the whole generation of the TSK fuzzy systems (2 hours approx. in a Pentium IV processor) was the 80% less than the amount needed for generating the original fuzzy revision subsystem. This timesaving operation is motivated by the simplified fuzzy rule base employed by the greedy algorithm used to generate each one of the TSK fuzzy systems.

Another relevant circumstance derived from the adoption of the proposed schema was the increment in the explanatory strength of the justification generated by the final CBR system. Initially the feature vector describing a problem was composed of 56 attributes, the same as the fuzzy rule antecedents, now the system is able to produce an explanation based on only 12 main features with the same level of accuracy.

6 Conclusion and Further Work

This paper introduces a new reduction technique based on Rough Set theory that can be applied for improving a previous successful method that automates the revision stage of CBR systems.

Empirical studies show that this reduction technique allow us to obtain a more general knowledge of the model and gain a deeper insight into the logical structure of the system to be approximated. Employing the simplified fuzzy rule base as the starting point to generate the fuzzy revision subsystem proposed in [14], leads to a dramatic decrease of the time needed for this task while maintaining an equivalent generalisation accuracy.

These benefits are augmented with the simplicity of the new fuzzy rules used by the CBR system as explanation of the final adopted solution. In this way, it is interesting the definition of a formal measure in order to rate and compare the explanation strength of these fuzzy rules.

Due to the suitability showed by the Rough Set theory working together with other soft computing techniques, we are also interested in the development of new ways to put together this formalism with the existing techniques coded in the FSfRT platform.

References

1. Riesbeck, C.K., Schank, R.C.: Inside Case-Based Reasoning, Lawrence Erlbaum Associates, Hillsdale, NJ, US (1999)
2. Pal, S.K., Dilon, T.S., Yeung, D.S.: Soft Computing in Case Based Reasoning, Springer Verlag, London (2000)
3. Sankar, K.P., Simon, C.K.S: Foundations of Soft Case-Based Reasoning, Wiley-Interscience, Hoboken, New Jersey (2003)
4. Pawlak, Z.: Rough Sets. International Journal of Computer and Information Sciences, Vol. 11. (1982) 341–356
5. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
6. Pawlak, Z.: Rough sets: present state and the future. Foundations of Computing and Decision Sciences, Vol. 11 (3-4). (1993) 157–166
7. Ziarko, W.: Variable Precision Rough Set Model. Journal of Computer and System Sciences, Vol. 46. (1993) 39–59
8. Düntsch, I., Gediga, G.: Statistical evaluation of rough set dependency analysis. International Journal of Human-Computer Studies, Vol. 46. (1997) 589–604
9. Düntsch, I., Gediga, G.: Uncertainty measures of rough set prediction. Artificial Intelligence, Vol. 106. (1998) 77–107
10. Díaz, F., Corchado, J.M.: A method based on the Rough Set theory and the MDL principle to select relevant features. Proc. of the X CAEPIA - V TTIA, Vol. 1. (2003) 101–104
11. Rissanen, J.: Minimum description length principle. In Kotz, S. and Johnson, N. L. (eds.). Encyclopedia of Statistical Sciences. John Wiley and Sons, New York (1985) 523–527
12. Blum, A.L., Langley, P.: Selection of relevant features and examples in machine learning. Artificial Intelligence, Vol. 97. (1997) 245–271
13. Fdez-Riverola, F., Corchado, J.M.: FSfRT, Forecasting System for Red Tides. An Hybrid Autonomous AI Model. Applied Artificial Intelligence, Vol. 17 (10). (2003) 955–982
14. Fdez-Riverola, F., Corchado, J.M.: An automated CBR Revision method based on a set of β -TSK Fuzzy models. Proc. of the X CAEPIA - V TTIA, Vol. 1. (2003) 395–404
15. Fdez-Riverola, F.: Neuro-symbolic model for unsupervised forecasting of changing environments. Ph.D. diss., Dept. of Computer Science, Vigo University, Spain (2002)